



TITLE:

# Quad-Tree based Image Encoding Methods for Data-Adaptive Visual Feature Learning( Abstract\_要旨 )

AUTHOR(S):

Zhang, Cuicui

---

CITATION:

Zhang, Cuicui. Quad-Tree based Image Encoding Methods for Data-Adaptive Visual Feature Learning. 京都大学, 2015, 博士(情報学)

ISSUE DATE:

2015-03-23

URL:

<https://doi.org/10.14989/doctor.k19111>

RIGHT:

( 続紙 1 )

京都大学	博士（情報学）	氏名	張 翠翠（Zhang, Cuicui）
論文題目	Quad-Tree based Image Encoding Methods for Data-Adaptive Visual Feature Learning （データ適応型特徴学習のための四分木に基づく画像の構造的表現法）		
(論文内容の要旨)			
<p>This dissertation addresses the visual feature learning problem in computer vision, and proposes a set of data-adaptive feature learning algorithms which explicitly explore the image spatial structures using Quad-Tree based image encoding methods. Based on experiments using facial image datasets and real world video datasets, these algorithms have been proved to be able to extract effective features for classification tasks and moving objects from real world video data including 3D scene structures and camera motions. The dissertation consists of seven chapters.</p> <p>Chapter 1 first introduces the visual feature learning problem and goal of this dissertation. Then, the motivation is given by discussing several challenging issues in existing algorithms. Afterwards, the proposed methods towards these challenging issues are presented. Finally, the main contributions and the roadmap of this dissertation are summarized.</p> <p>Chapter 2 reviews problems in visual feature learning. It starts from the widely accepted concept: low-level, mid-level, and high-level features, and a discussion on corresponding holistic/local feature learning algorithms. Then, a survey is followed on the recent developed hierarchical learning algorithms. Finally, it introduces another visual feature learning technique: ensemble learning, which can be applied to the Small Sample Size problem.</p> <p>Chapter 3 addresses the issue of exploring the image spatial structure for feature learning, and proposes a data-adaptive block decomposition by a multi-threshold Quad-Tree. It performs Quad-Tree partition on a template image of an image dataset instead of raw images to make the Quad-Tree partition adaptive to the image dataset. Then, the resulting partition depicts the feature distribution on images in the dataset for effective low-level feature learning. The algorithm is evaluated with four large face databases for recognition, and performs comparably with the-state-of-arts.</p> <p>Chapter 4 addresses the issue of exploring the image hierarchical structure to learn the high-level features from the local feature composition. It utilizes a tree structure with varied edge weights to depict a fact that the different subregions contribute differently in the feature learning. This hierarchical structure is defined as a top-down Quad-Tree structure, and utilized to learn high level features in a bottom-up manner. It is more data-adaptive compared with other image pyramid models. Moreover, varied weight assignment strategy makes the global</p>			

feature learning more accurately. Experimental results show the effectiveness and efficiency of this method on complex face databases.

Chapter 5 studies the Small Sample Size problem; many practical problems have no sufficient training data. The proposed method first augments the small training data using random matrix, and then encodes the feature from the enlarged training data by Quad-Trees. These features are employed to integrate diverse base classifiers in an ensemble scheme for robust classification based on the 0-1 Knapsack problem formulation. The proposed method shows a better performance on the SSS face recognition compared with the state-of-the-arts.

Chapter 6 investigates the problem of extracting moving real world objects from videos captured by a moving camera. The proposed method first transforms the 2D motion flow into a 3D potential surface using Helmholtz-Hodge decomposition to formulate the background motion, and then utilizes the Quad-Tree to label each foreground motions and depth discontinuities for motion segmentation. The experiments with four benchmark video databases demonstrate the proposed method outperforms the latest methods.

Chapter 7 summarizes the proposed Quad-Tree based image encoding methods for data-adaptive visual feature learning, and lists the possible directions for future work.

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 words で作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

## (論文審査の結果の要旨)

本論文は、顔画像認識のための視覚特徴学習および動画像からの運動対象検出に対し、4分木を用いたデータ適応型のアルゴリズムを複数提案し、それらの有効性を多様な顔画像データベースおよび3次元構造・運動を含んだ実世界動画像を用いた実験によって実証したもので、得られた主な成果は次の通りである。

1. 顔画像認識に関しては多くの先行研究が存在するが、本研究では、認識に有効な特徴は空間的に一様に分布するのではなく、顔の持つ空間構造に応じて有効な特徴の位置、大きさが変化するため、顔画像データベースから自動的に有効な特徴を学習することが重要であることを指摘し、4分木を用いた適応的な特徴学習アルゴリズムを考案し、多様な顔画像データベースを用いた実験によって、従来法より優れた認識性能を持つことを示した。
2. 一般の画像認識では、学習に利用できるデータ量が限られているという、小規模サンプル集合 (small sample size: SSS) 問題がある。本研究では、小規模顔画像データベースから、4分木を用いて顔画像の持つ空間的特徴分布を求め、4分木の構造に応じて特徴にランダムな変動を与えることによって、学習データ数の拡張を図るアルゴリズムを考案し、その有効性を実験によって実証した。
3. 一般に画像認識では、画像から抽出された多数の特徴をどのように組み合わせるのか、異なった特性を持つ認識アルゴリズムをどのように組み合わせるのかといった、特徴選択・統合、認識アルゴリズム統合が問題となる。本研究では、これらの問題を0-1ナップサック問題として定式化し、上記1、2における認識アルゴリズムに適用し、その有効性を実証した。
4. 3次元構造・運動を含んだ実世界動画像から運動対象を検出するには、カメラ運動と対象運動とを分離することが必要となる。本研究では、動画像から抽出された2次元オブティカルフローをHelmholtz-Hodge decompositionによって3次元のポテンシャル面に変換し、ポテンシャル面の空間的变化を4分木によって構造的に表現し、各部分空間の示す運動情報を基にカメラ運動成分と運動対象成分を分離するアルゴリズムを提案し、様々な実世界映像から運動対象が精度よく検出できることを示した。

以上本論文は、顔画像認識のための視覚特徴学習に対し、4分木を用いたデータ適応型の視覚特徴学習アルゴリズムおよび0-1ナップサック問題に基づいた特徴統合、認識アルゴリズム統合法を提案し、それらの有効性を多様な顔画像データベースを用いた実験によって実証するとともに、3次元構造・運動を含んだ実世界動画像の示す複雑なオブティカルフローを4分木を用いて適応的に分割、分析することによって運動対象を精度よく検出できることを示したもので、学術上、實際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成27年2月23日論文内容とそれに関連した口頭試問を行った結果合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。

更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。

要旨公開可能日： 年 月 日以降